SAND2013-7969P

The Zoltan Toolkits: Parallel Partitioning, Load Balancing, Coloring and Ordering

Karen Devine, **Erik Boman, Siva Rajamanickam, Vitus Leung**

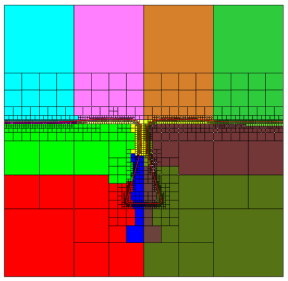
Chevron Visit; September 13, 2013

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000

Zoltan Toolkits of Parallel Combinatorial Algorithms

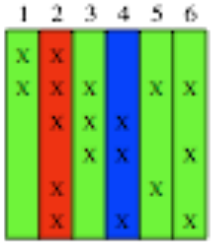
Sandia National Laboratories

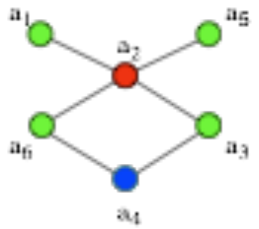
Dynamic Load Balancing



Assign data and work to processes in a way that minimizes application execution time.

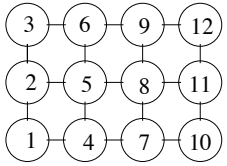
Graph Coloring

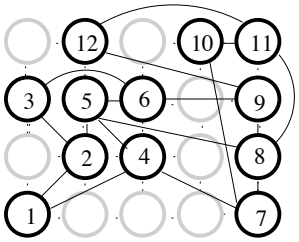




Label vertices to identify sets of independent operations

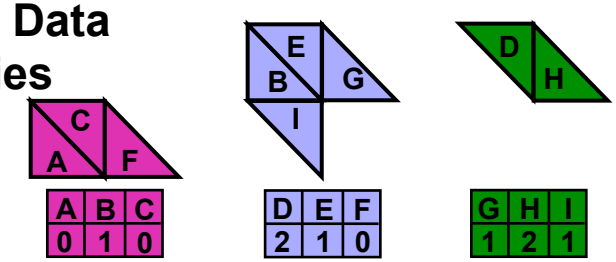
Task Mapping





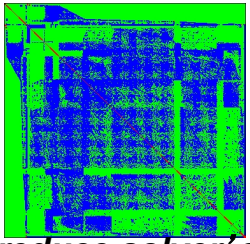
Map ranks to allocated nodes and cores to reduce communication time and congestion.

Distributed Data Directories



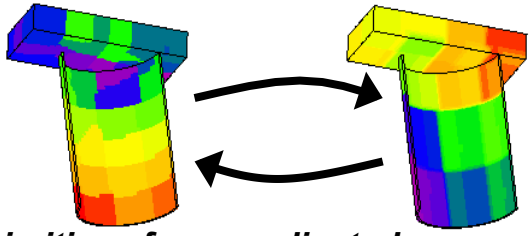
Efficiently locate needed off-processor data.

Graph Ordering



Permute matrix to reduce solver's memory use and/or data movement.

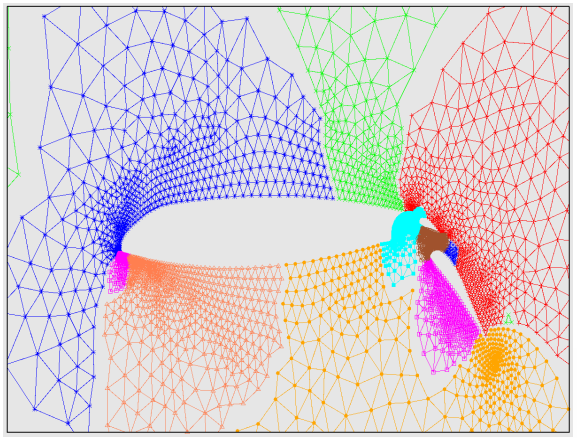
Unstructured Communication

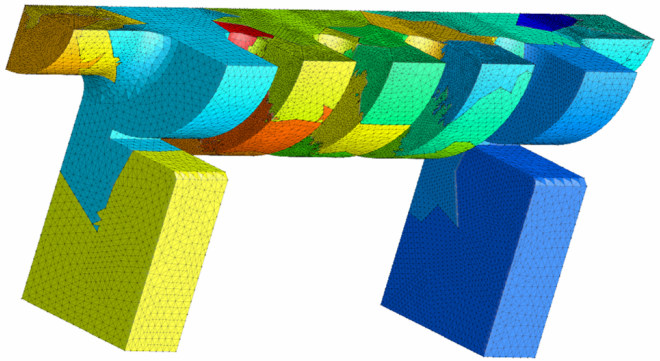


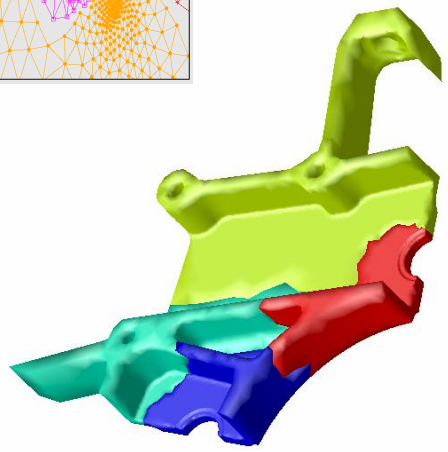
Provide simple primitives for complicated communication patterns.

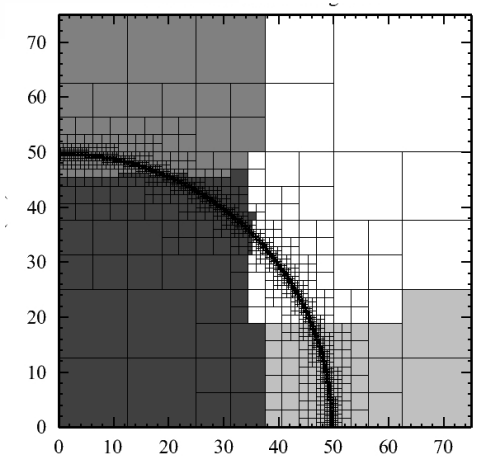
Partitioning and Load Balancing

- Assignment of application data to processors for parallel computation.
- Applied to grid points, elements, matrix rows, particles, ….



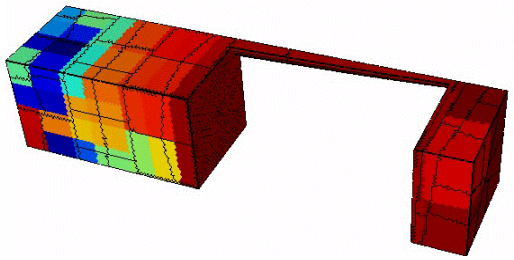


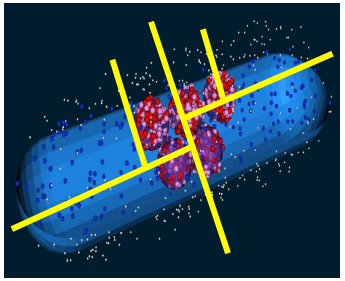




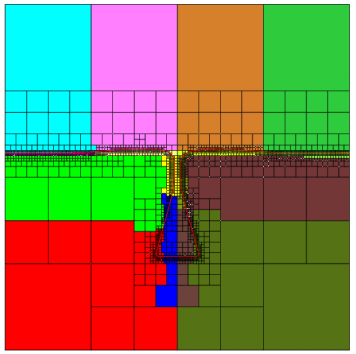
Zoltan is used in many different applications.

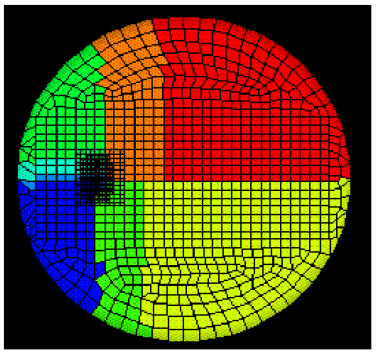
Sandia National Laboratories



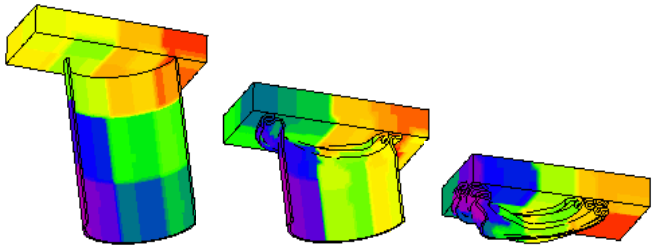


Balancing and maintaining locality of particles in neutron generator and chemical cell simulations.

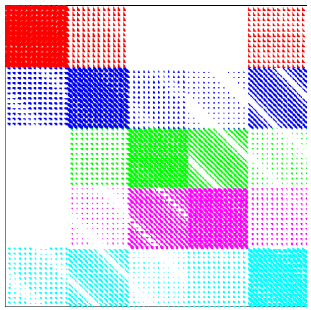


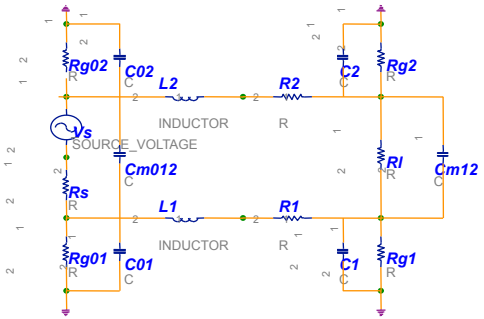


Balancing adaptively refined meshes for safety analysis, shock physics, chemical vapor deposition, CFD, etc.,



Maintaining geometric locality of surfaces for contact detection and crash simulations





Balancing linear systems arising in electronic circuit simulations

Redistributing matrices in multilevel preconditioners

Zoltan Toolkit: Suite of Partitioners

Sandia National Laboratories

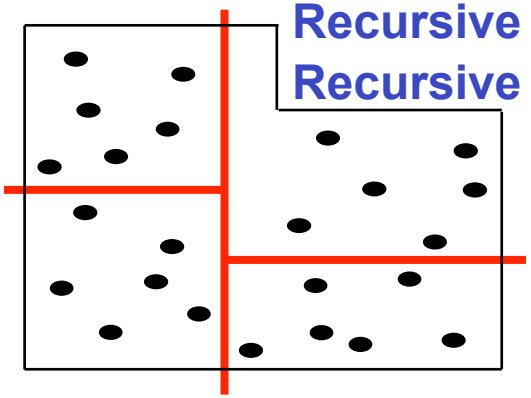
- No single partitioner works best for all applications.
  - Trade-offs:
    - Quality vs. speed.
    - Geometric locality vs. data dependencies.
    - High-data movement costs vs. tolerance for remapping.
- Application developers may not know which partitioner is best for application.

- Zoltan contains suite of partitioning methods.
  - Application changes only one parameter to switch methods.
    - Zoltan_Set_Param(zz, "LB_METHOD", "new_method_name");
  - Allows experimentation/comparisons to find most effective partitioner for application.

Partitioning Algorithms in the Zoltan Toolkit

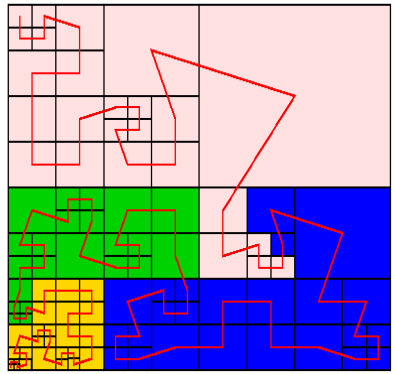
Sandia National Laboratories

Geometric (coordinate-based) methods



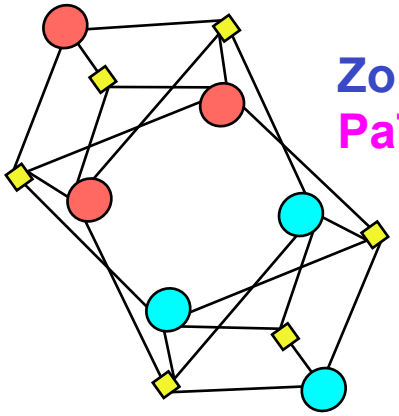
Recursive Coordinate Bisection (Berger, Bokhari)

Recursive Inertial Bisection (Taylor, Nour-Omid)

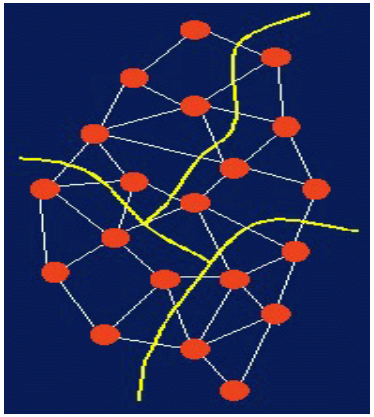


Space Filling Curve Partitioning
(Warren&Salmon, et al.)

Combinatorial (topology-based) methods



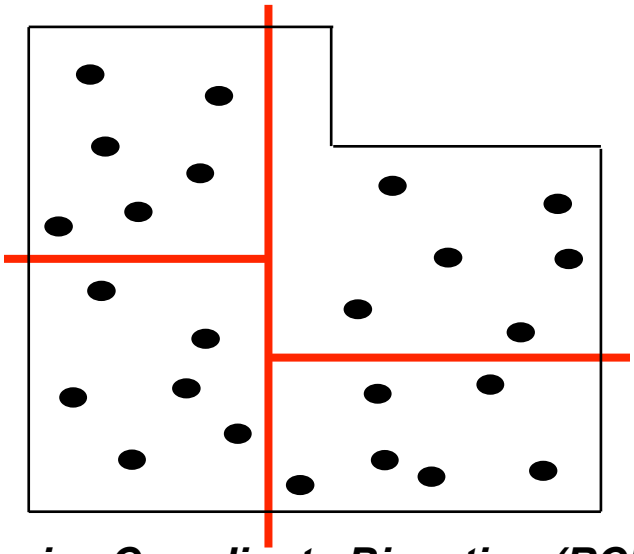
Zoltan Hypergraph Partitioning (PHG)
PaToH (Catalyurek & Aykanat)



Zoltan Graph Partitioning (PHG)
ParMETIS (Karypis, et al.)
PT-Scotch (Pellegrini, et al.)

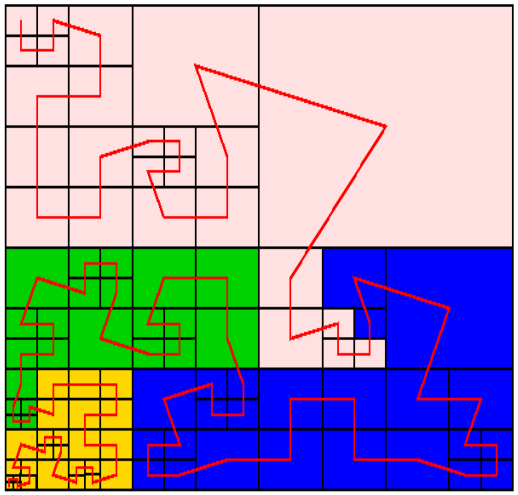
Geometric Partitioning

- **Partition geometric coordinates to maintain locality.**
  - **Create parts containing physically close data.**



Recursive Coordinate Bisection (RCB)
Berger & Bokhari, 1987
Recursive Inertial Bisection (RIB)
Taylor, Nour-Omid, 1994

Compute cutting planes that recursively divide work



Space Filling Curve Partitioning (HSFC)
Warren & Salmon, 1993;
Pilkington & Baden, 1994; Patra & Oden, 1995

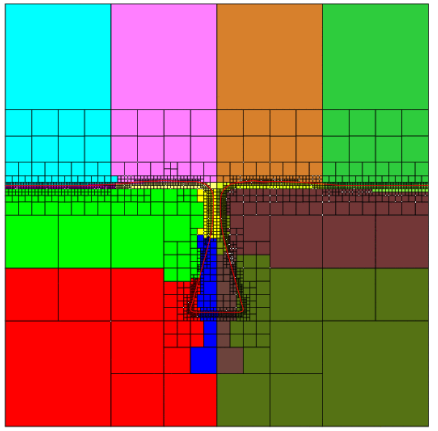
Partition linear ordering of data given by a space-filling curve

Sandia National Laboratories

Geometric Partitioners

- **Advantages:**
  - Conceptually simple; fast and inexpensive.
  - Effective when connectivity info is not available (e.g., in particle methods).
  - All processors can inexpensively know entire decomposition.
  - RCB: Regular subdomains useful in structured or unstructured meshes.
  - SFC: Linear ordering may improve cache performance.

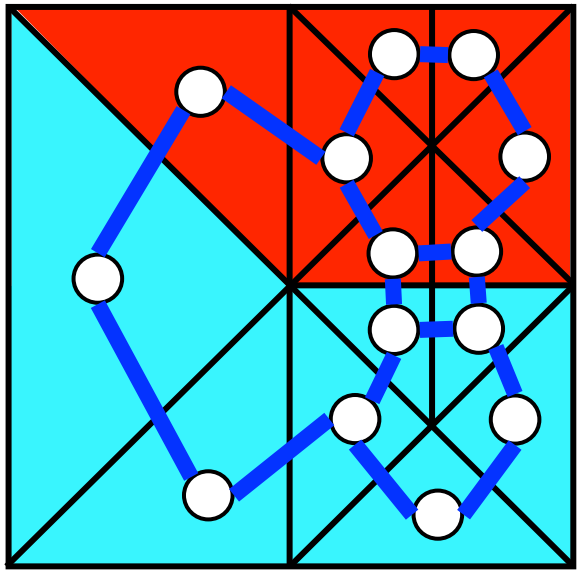
- **Disadvantages:**
  - No explicit control of communication costs.
  - Can generate disconnected subdomains for complex geometries.
  - Geometric coordinates needed.



Sandia National Laboratories

Topology-based Partitioners

- **Goal: Balance work while minimizing data dependencies between parts.**
  - **Represent data with vertices of graph/hypergraph**
  - **Represent dependencies with graph/hypergraph edges**
  - **Partition so that parts have equal vertex weights and weights of (hyper)edges cut by part boundaries is small.**

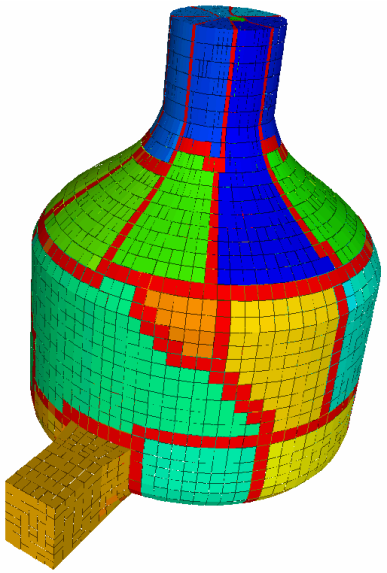


Kernighan, Lin, Schweikert, Fiduccia, Mattheyes, Simon, Hendrickson, Leland, Kumar, Karypis, Alpert, Kahng, Hauck, Borriello, Çatalyürek, Aykanat, et al.

Topology-based Partitioning: Advantages and Disadvantages

Sandia National Laboratories

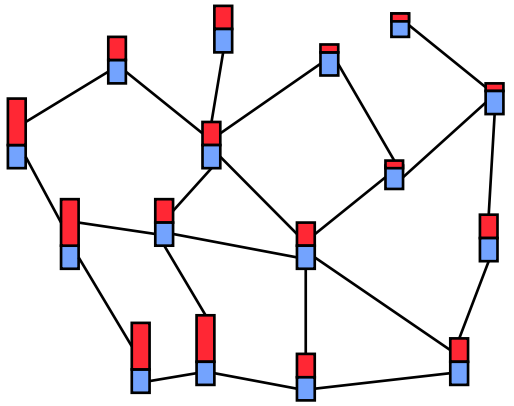
- **Advantages:**
  - **Explicit control of communication volume gives higher partition quality (i.e., lower total communication volume) than geometric methods.**
  - **Excellent software available.**
    - **Serial graph:** Chaco (SNL), Jostle (U. Greenwich), METIS (U. Minn.), Scotch (U. Bordeaux)
    - **Serial hypergraph:** Mondriaan (Utrecht U.), PaToH (Ohio St. U.)
    - **Parallel graph:** Zoltan (SNL), ParMETIS (U. Minn.), PT-Scotch (LaBRI/INRIA), PJostle (U. Greenwich)
    - **Parallel hypergraph:** Zoltan (SNL)



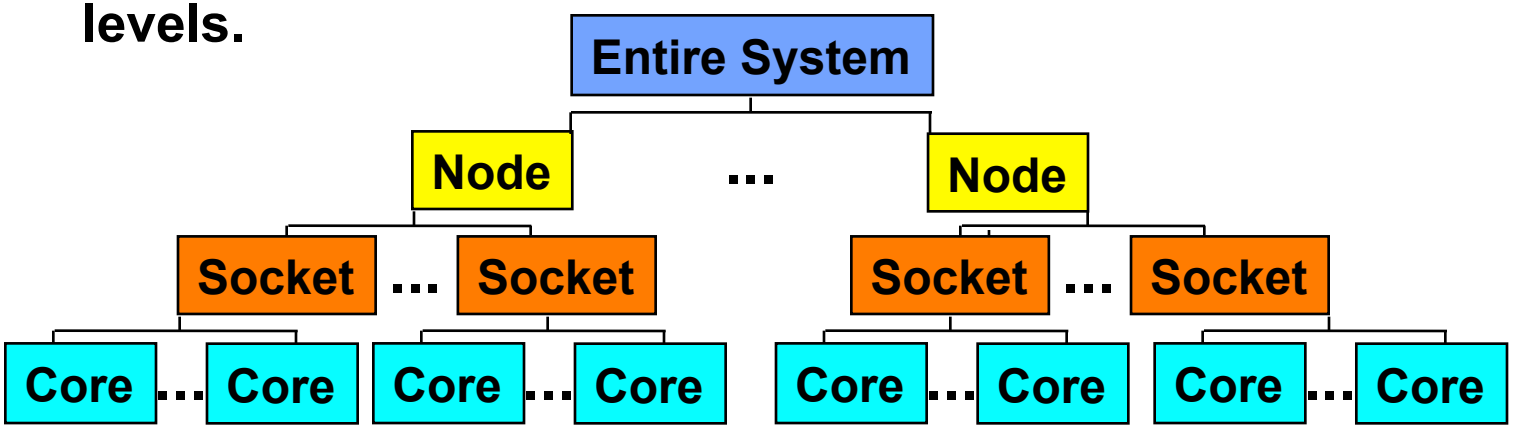
- **Disadvantages:**
  - **More expensive than geometric methods.**
  - **Explicit dependence info is required.**

Other Partitioning Features

- **Multiple constraints or objectives**
  - Compute a single partition that is good with respect to multiple factors.
    - Balance both computation and memory
    - Balance multi-phase simulations
  - Available in RCB, RIB and ParMETIS graph.



- **Hierarchical partitioning for multicore**
  - Partition at each level of machine hierarchy
  - Allows different partitioners at different architecture levels.



Sandia National Laboratories

Interfaces to Zoltan and Zoltan2

- **Native Zoltan interface: C, C++ and F90**
  - **Callback functions give application data to Zoltan**
- **Isorropia package in Trilinos**
  - **Epetra matrix/vector interface to Zoltan**
- **New user-data adapter interfaces in Zoltan2**
  - **Matrices, meshes, graphs, coordinates**
  - **Epetra and Tpetra supported through Xpetra**

Current Research and Development

Sandia National Laboratories

- **Partitioning for scale-free graphs**
  - New method combines 2D data distribution + 1D graph partitioning
  - Reduced SpMV time by 33% (on average)
  - SC13 paper: “Scalable Matrix Computations on Large Scale-Free Graphs Using 2D Graph Partitioning.” (Boman, Devine, Rajamanickam)
- **Architecture-aware partitioning and task mapping**
  - Incorporate static and real time system information into decisions on work distribution
  - Intelligent task mapping based on geometric partitioning reduced communication time of MiniGhost miniapp by 8-40%.

Sandia National Laboratories

For More Information...

- Zoltan Home Page
  - http://www.cs.sandia.gov/Zoltan
  - User's and Developer's Guides
  - Tutorial: "Getting Started with Zoltan: A Short Tutorial"

- Email
  - kddevin@sandia.gov